Encoding and Decoding Models

Francisco Pereira
Machine Learning Team
National Institute of Mental Health

"All models are wrong, but some are useful."

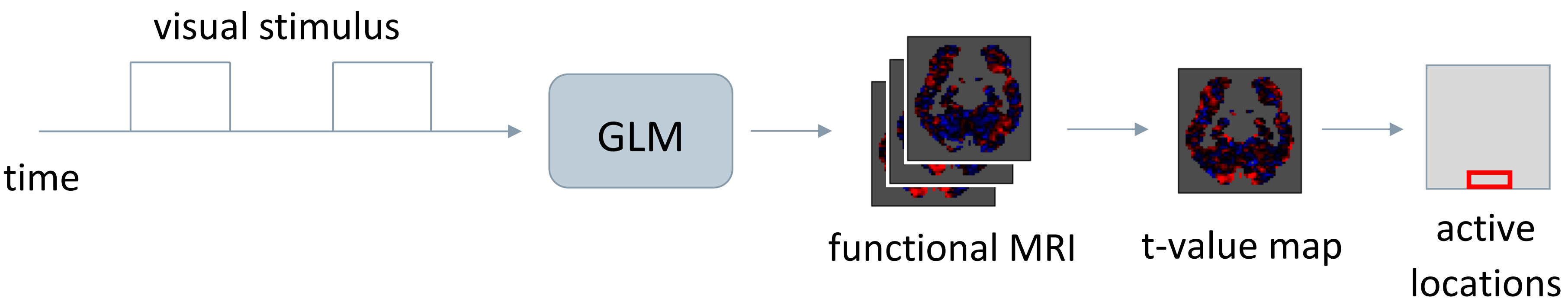
George Box

”I hope you paid attention during Martin’s talk...”

Francisco Pereira

overview

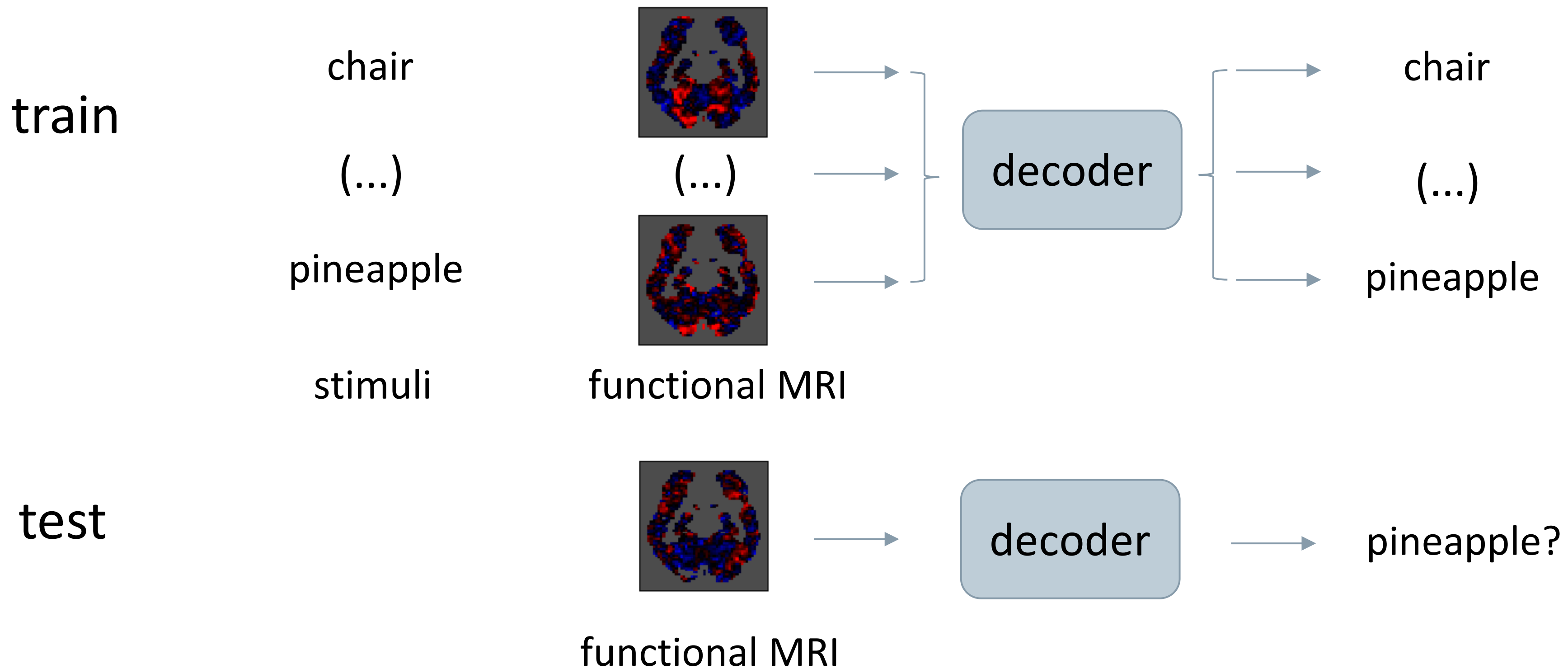
activation



- GLM: general linear model
- mass univariate: regression model of each voxel from the stimulus
- refinements: nuisance regressors, graded responses, ...
- good GLM: explains voxel behavior

overview

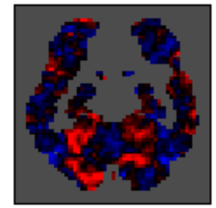
decoding



- decoder: classifier, regression model, ...
- multivariate: input is pattern of activation over many voxels
- good decoder: accurate at picking correct stimulus

tools and questions

stimulus
(task)



GLM

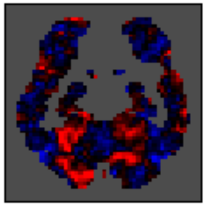
“is there a brain location that responds to the stimulus?“

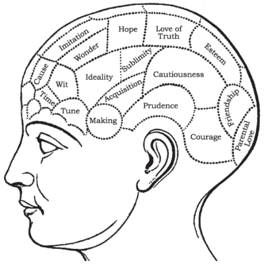
decoder

“is there information about the stimulus
in the pattern of activation?“

tools and questions

stimulus
(task)





GLM

“is there a brain location that responds to the stimulus?“

decoder

“is there information about the stimulus
in the pattern of activation?“

... and the questions we care about

- what does brain area X do?
- is brain area X used in task Y?
- if a subject is doing Y, what does their brain represent?
- where are representations invariant to Z?
- does everyone with disease W have altered connectivity to X?
- ...

using GLMs to answer questions

”Is region X used in task Y?”

- careful choice of control condition(s)

  (e.g. if studying sentences, control with nonwords, scrambled words,...)

- using a meta-analysis to perform reverse inference [Poldrack 2011]
  - “how likely is task Y given region X?”
  - activation databases
    - BrainMap, NeuroSynth, NeuroVault, ...
    - activation locations, statistical maps, ...
    - fixed ontologies of neural function vs terms extracted from paper text

using decoders to answer questions

- restrict voxels considered in space or time
- select voxels by their behavior
- use decoders as sensors of cognitive states
- ...

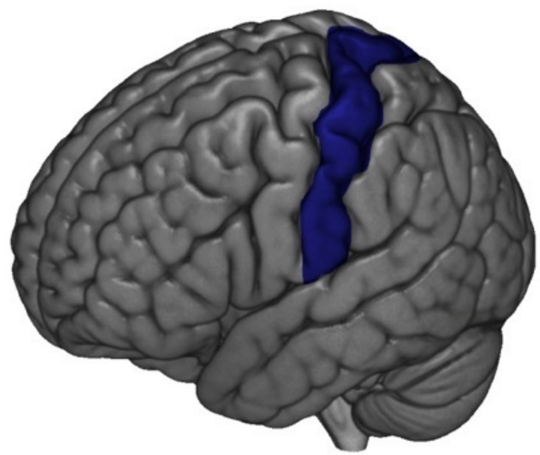
using decoders to answer questions

restrict voxels considered in space or in time

whole-brain

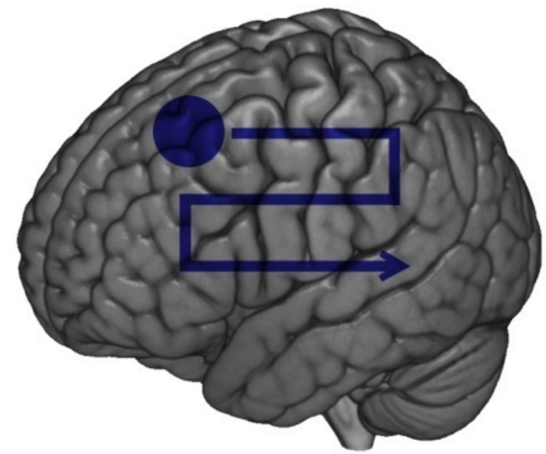
one accuracy result

region of interest

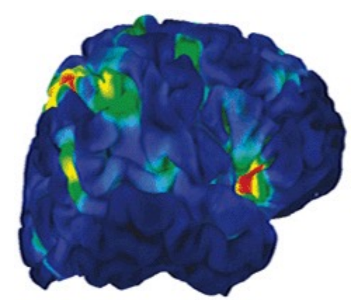


one accuracy result per ROI

searchlight



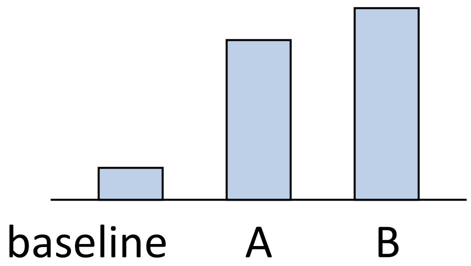
one accuracy result per searchlight



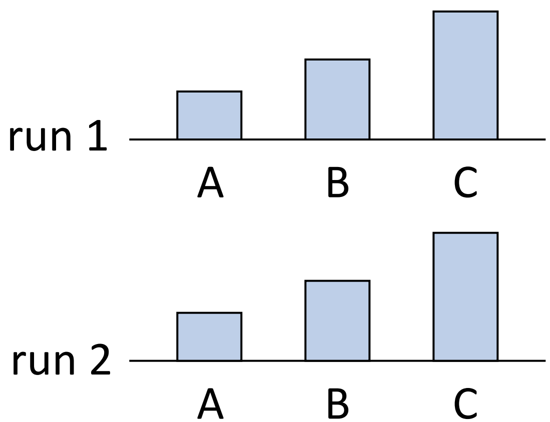
[adapted from Martin Hebart]

using decoders to answer questions

select input voxels by their behavior



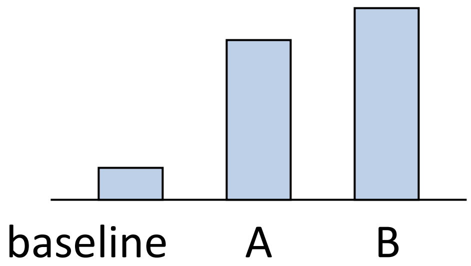
active during task



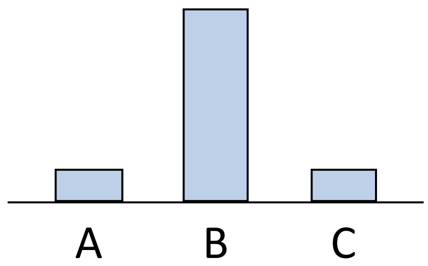
responds consistently
to each condition

using decoders to answer questions

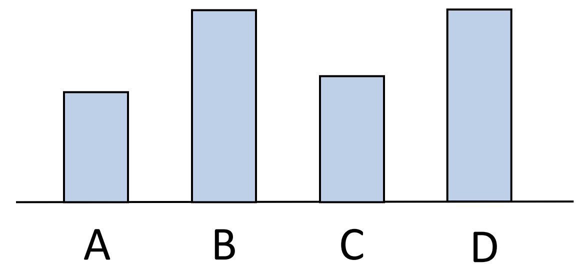
select input voxels by their behavior



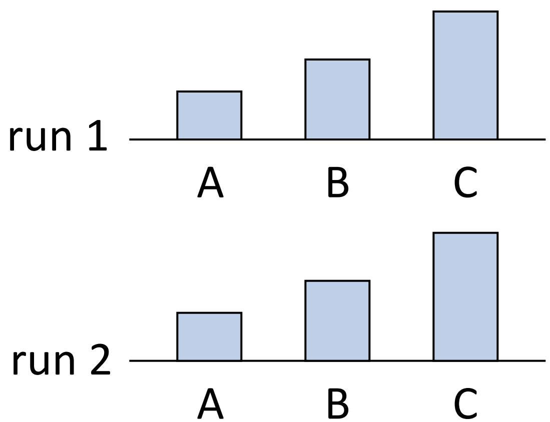
active during task



selective for condition



responds differently
to some conditions



responds consistently
to each condition

- heightened danger of circularity...

[Pereira et al 2009]

[Kriegeskorte et al 2009]

using decoders to answer questions

decoders as virtual sensors of cognitive states in time...

train decoders of faces / locations / objects on study phase fMRI

face decoder

location decoder

object decoder

[Polyn et al 2005]

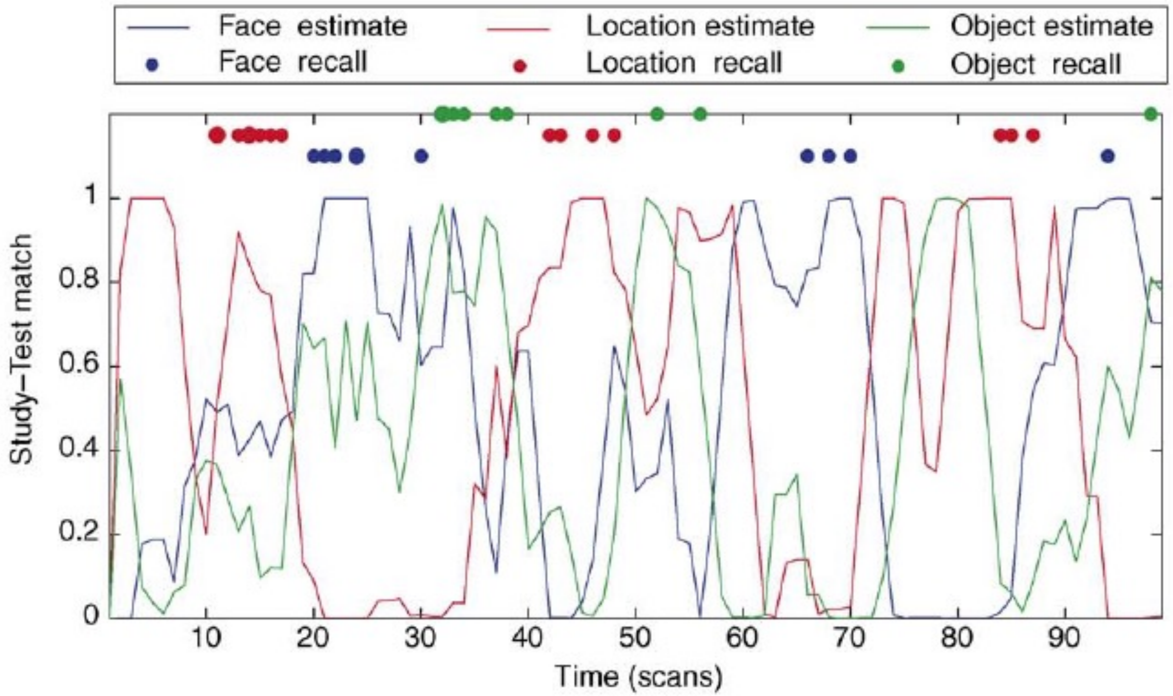
using decoders to answer questions

decoders as virtual sensors of cognitive states in time...

train decoders of faces / locations / objects on study phase fMRI

face decoder | location decoder | object decoder

apply decoders to detect category reinstatement during free-recall fMRI

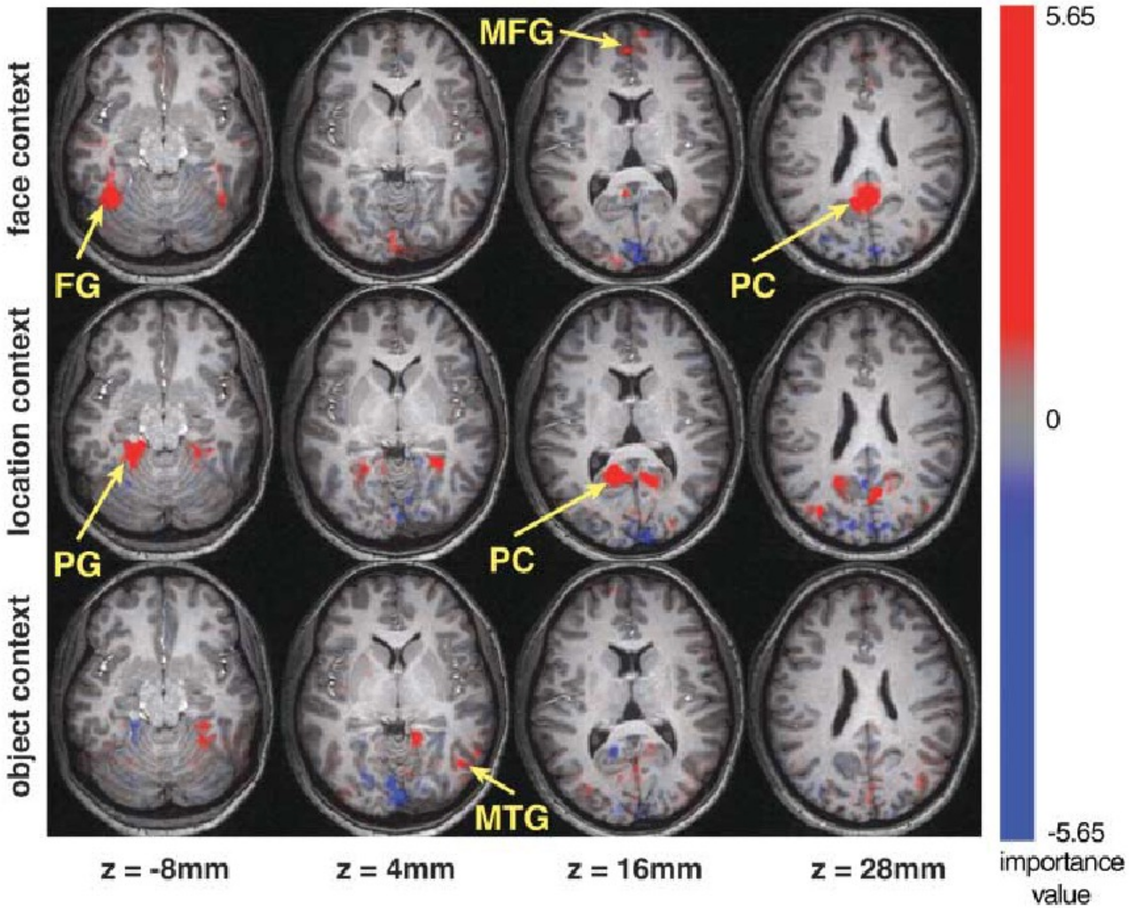


[Polyn et al 2005]

using decoders to answer questions

... also shed light on spatial distribution of information

voxels with the most impact on each decoding estimate



face decoder

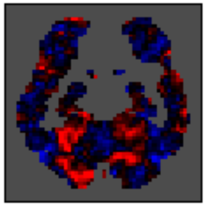
location decoder

object decoder

[Polyn et al 2005]

inference seems rather indirect...

stimulus
(task)



GLM inference

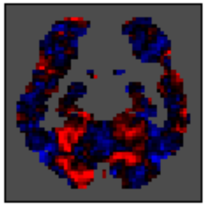
driven by task contrasts, prior studies

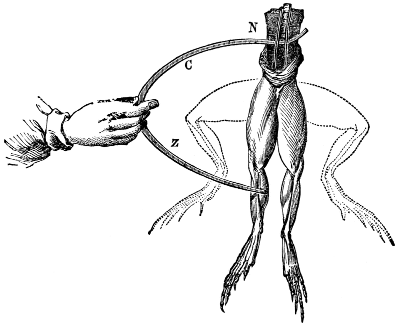
decoder inference

driven by feature choices,
and dissection of decoders

inference seems rather indirect...

stimulus
(task)





GLM inference

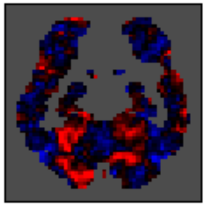
driven by task contrasts, prior studies

decoder inference

driven by feature choices,
and dissection of decoders

... but it doesn‘t have to be!

stimulus
(task)

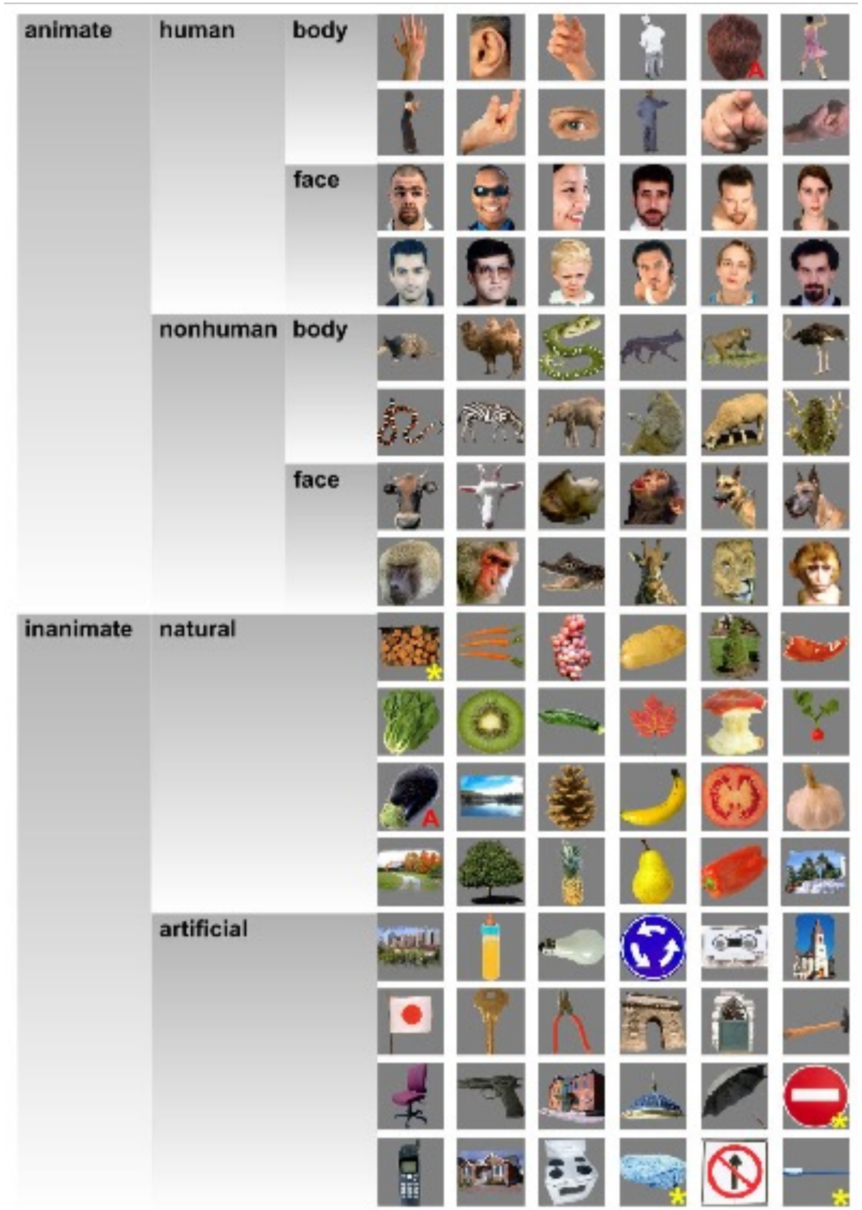


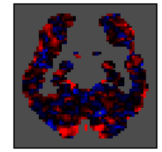
representation

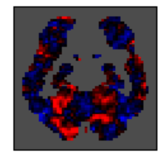
what is represented in the brain as a task is performed?

- known or constrained by behavioral or animal experiments
- mathematical or computational models
- hypothesized
- learned elsewhere (text corpora, image database, ...)
- ...

representational similarity analysis

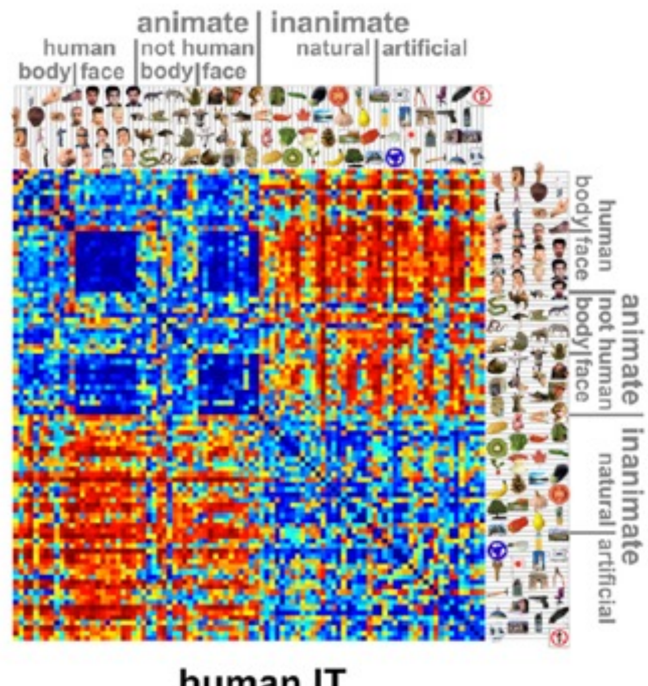




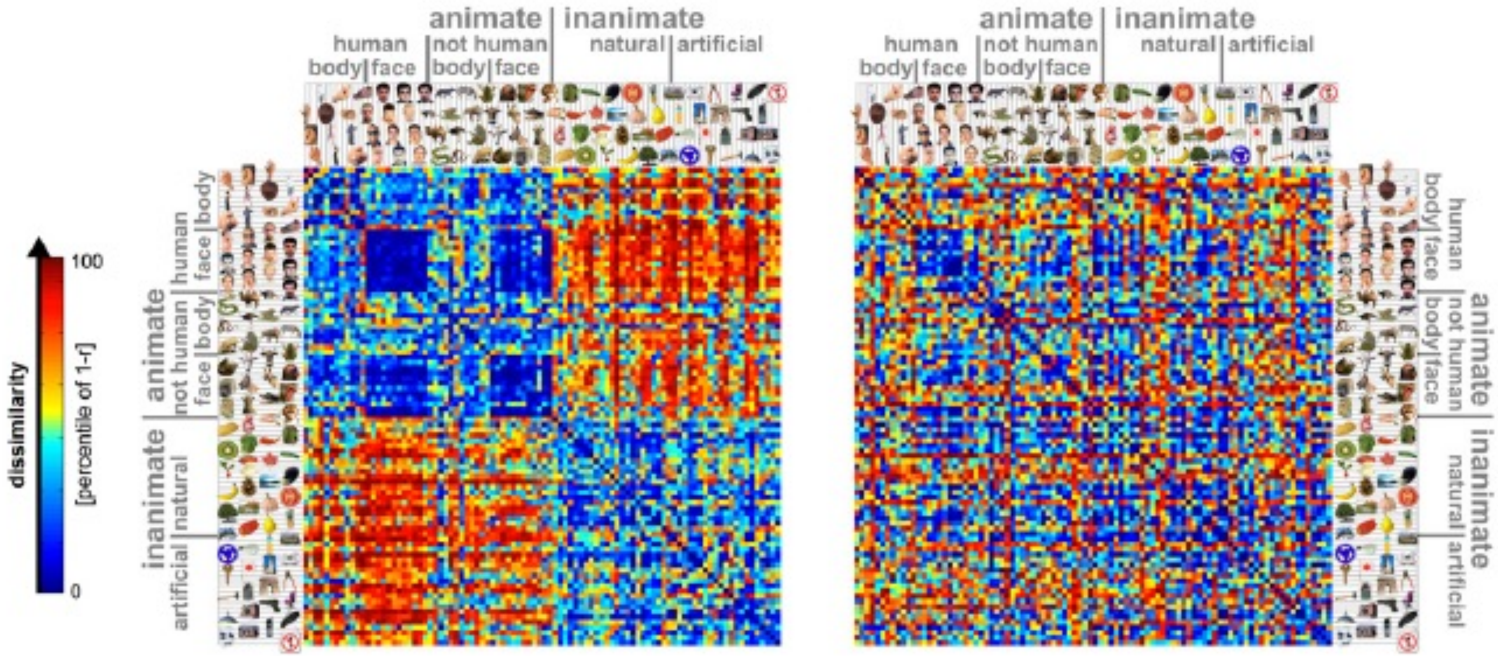


[Kriegeskorte, 2008]

calculate similarity of activation patterns



representational similarity analysis



human IT

human early visual cortex

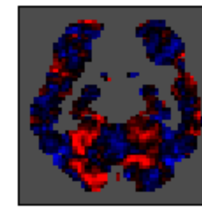
- similarity structure between activation patterns differs by location
- different contrasts allow inference about what is represented

encoding and decoding models

encoding model

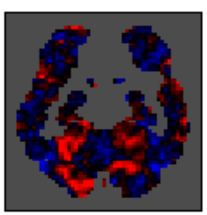
“does my representation predict activation?“

stimulus (task) → representation → mapping →



decoding model

“can I infer my representation (and stimulus) from activation?“



→ mapping → representation → stimulus (task)

case study 1 (encoding)

Predicting Human Brain Activity Associated with the Meanings of Nouns

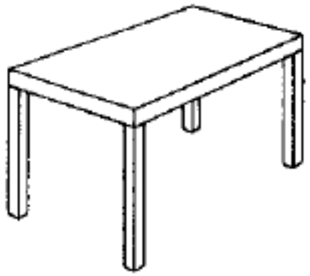
Tom M. Mitchell,[1*] Svetlana V. Shinkareva,[2] Andrew Carlson,[1] Kai-Min Chang,[3,4] Vicente L. Malave,[5] Robert A. Mason,[3] Marcel Adam Just[3]

[Science, 2008]

design

stimulus in each trial

(3 sec + 8 sec fixation)

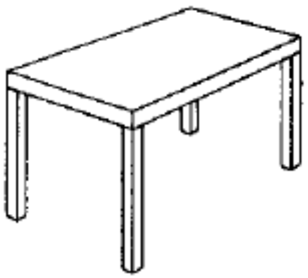


Table

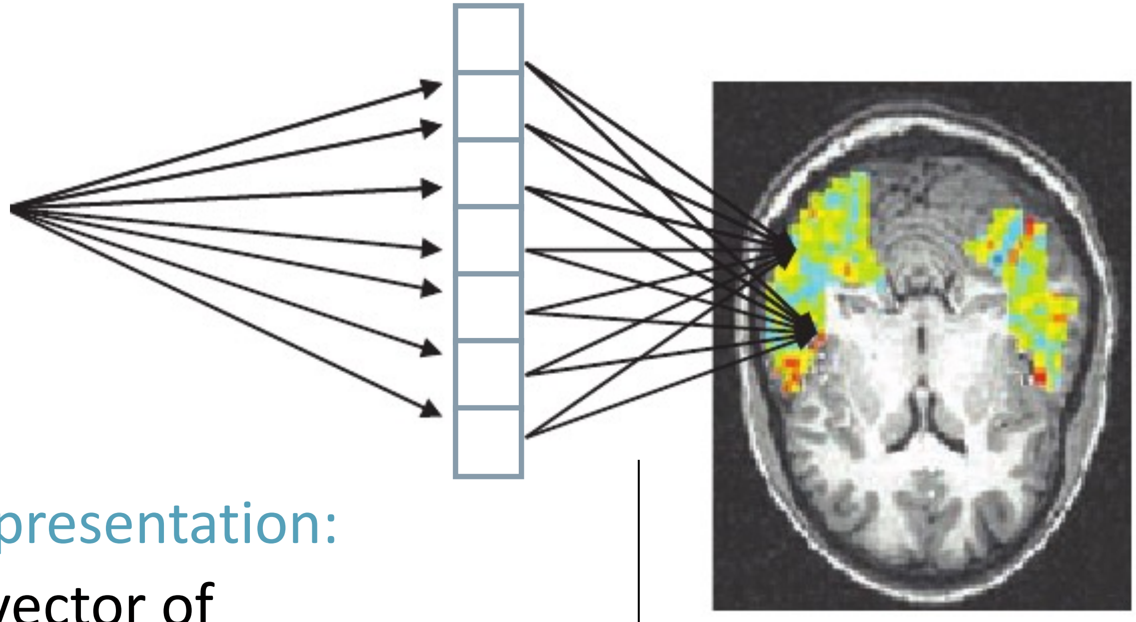
60 different words (12 categories x 5 exemplars)

BODY PARTS	leg	arm	eye	foot	hand
FURNITURE	chair	table	bed	desk	dresser
VEHICLES	car	airplane	train	truck	bicycle
ANIMALS	horse	dog	bear	cow	cat
KITCHEN UTENSILS	glass	knife	bottle	cup	spoon
TOOLS	chisel	hammer	screwdriver	pliers	saw
BUILDINGS	apartment	barn	house	church	igloo
PART OF A BUILDING	window	door	chimney	closet	arch
CLOTHING	coat	dress	shirt	skirt	pants
INSECTS	fly	ant	bee	butterfly	beetle
VEGETABLES	lettuce	tomato	carrot	corn	celery
MAN MADE OBJECTS	refrigerator	key	telephone	watch	bell

model



Table



representation:

a vector of
semantic features

average activation
during 3 seconds

mapping:

each voxel is a linear
combination of
semantic features

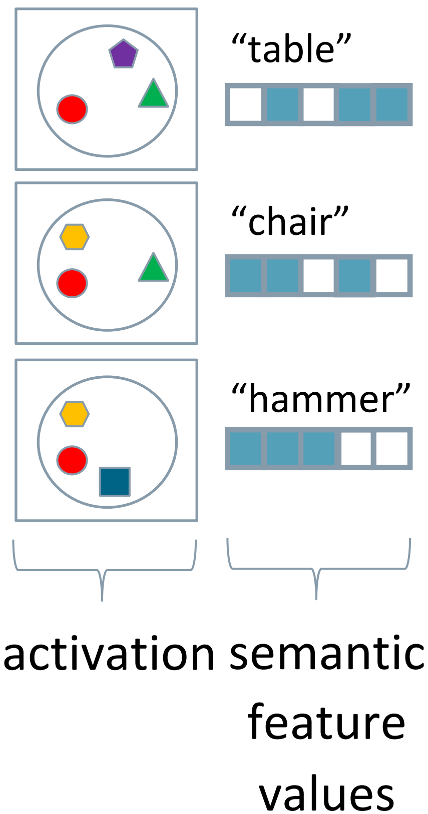
representation

- goal: represent different aspects of meaning of a word

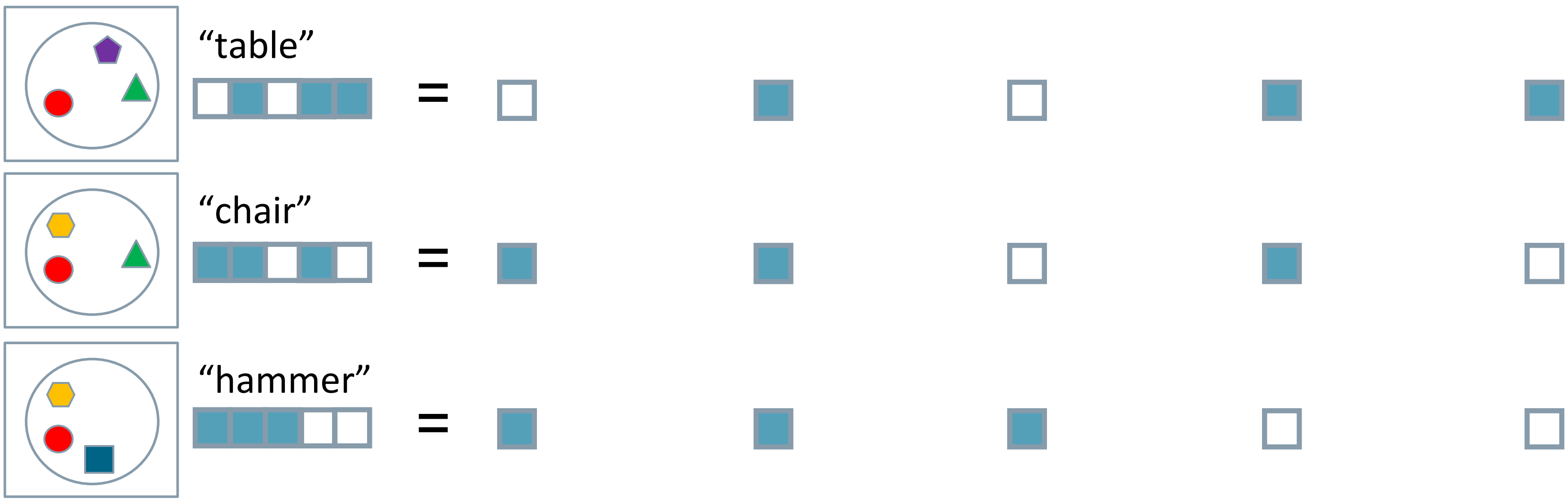
- 25 verbs used as proxies:
  - sensory: see, hear, listen, taste, touch, smell, fear, ...
  - motor: rub, lift, run, push, move, say, eat, ...
  - other: fill, open, ride, approach, drive, enter, ...
- 25 feature values:
  - co-occurrence of each word with 25 verbs, in a large text corpus

- e.g. “airplane” (0.87, ride, 0.29, see, 0.17, near, 0.08, run, ...)

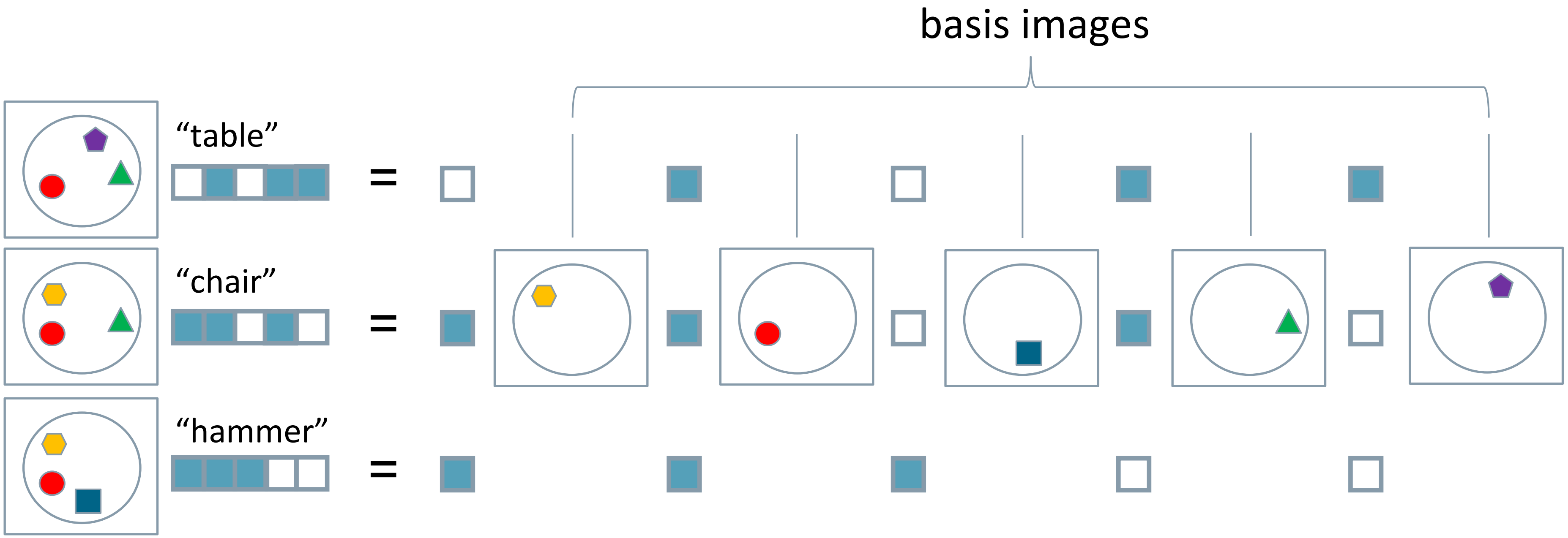
mapping



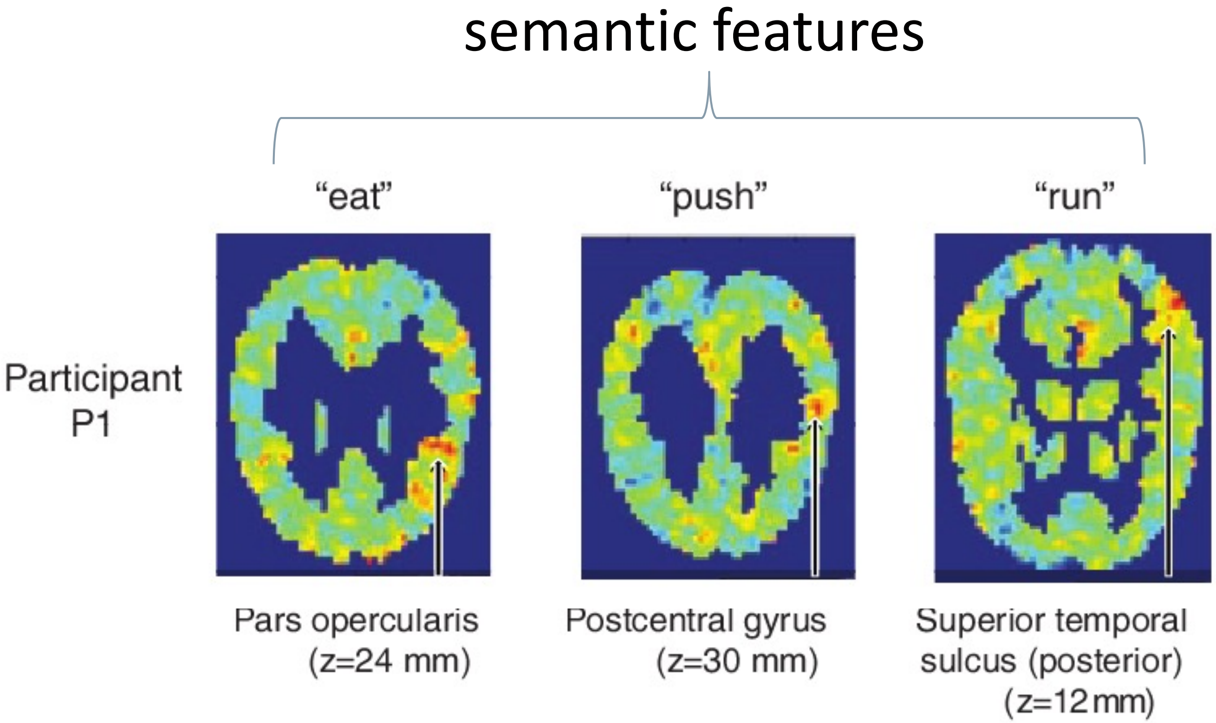
mapping



mapping



mapping

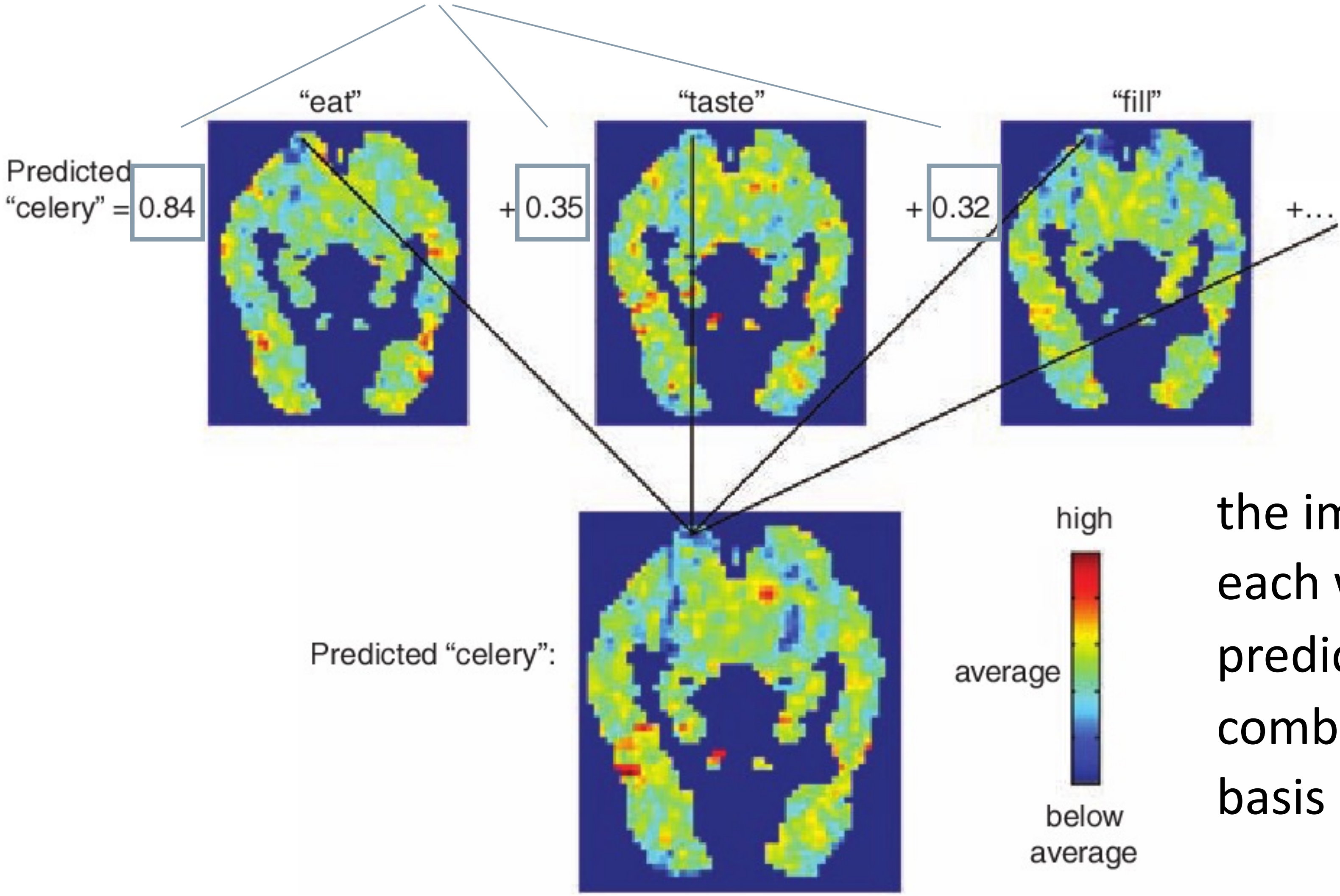


basis images capture the presence of each semantic feature across the brain

30

mapping

semantic feature values for “celery”



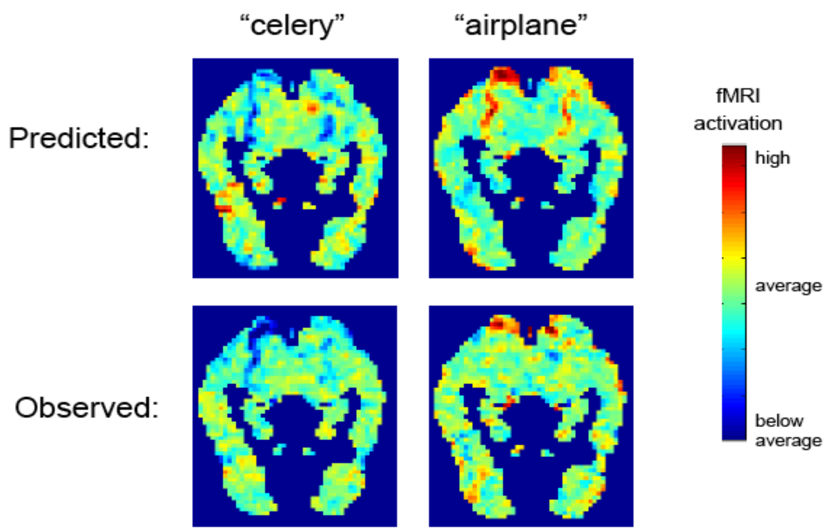
the image for each word is predicted as a combination of basis images

31

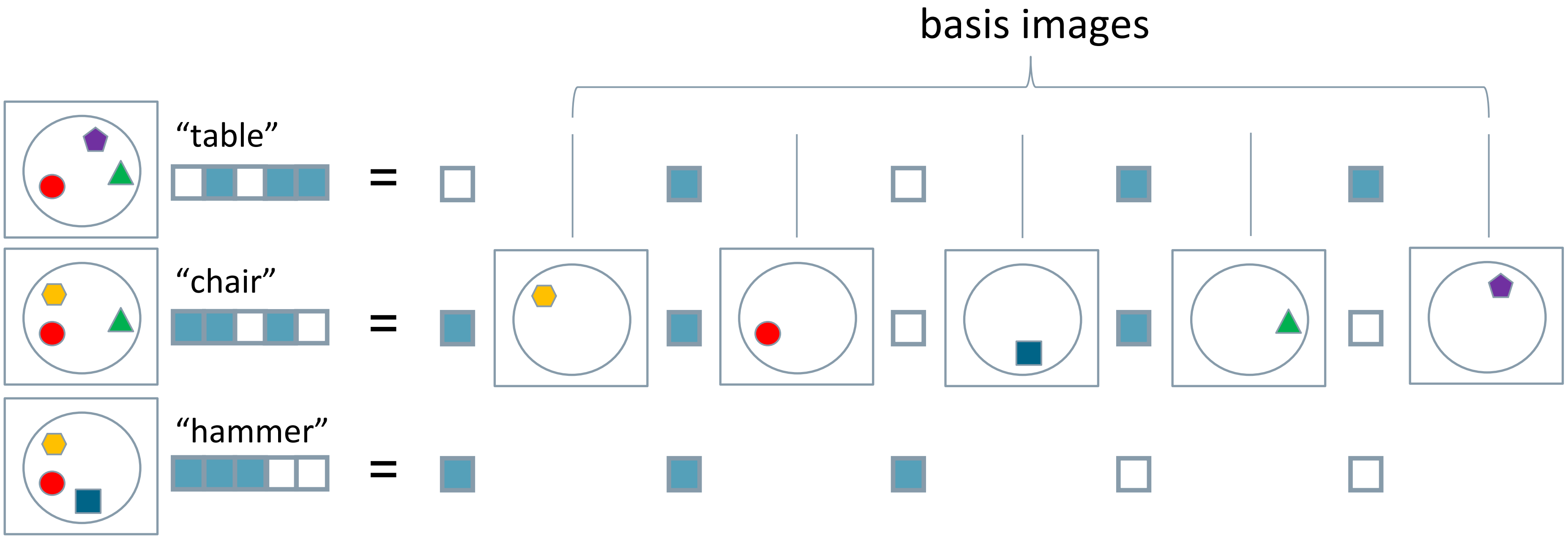
evaluation

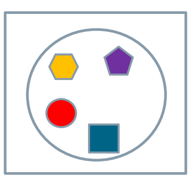
- learn basis images from 58 of the 60 words
- predict images for 2 left-out test words ("celery" and "airplane"), from their semantic feature values + basis images
- correct prediction if predicted can be matched to observed

(average accuracy across subjects 72%)



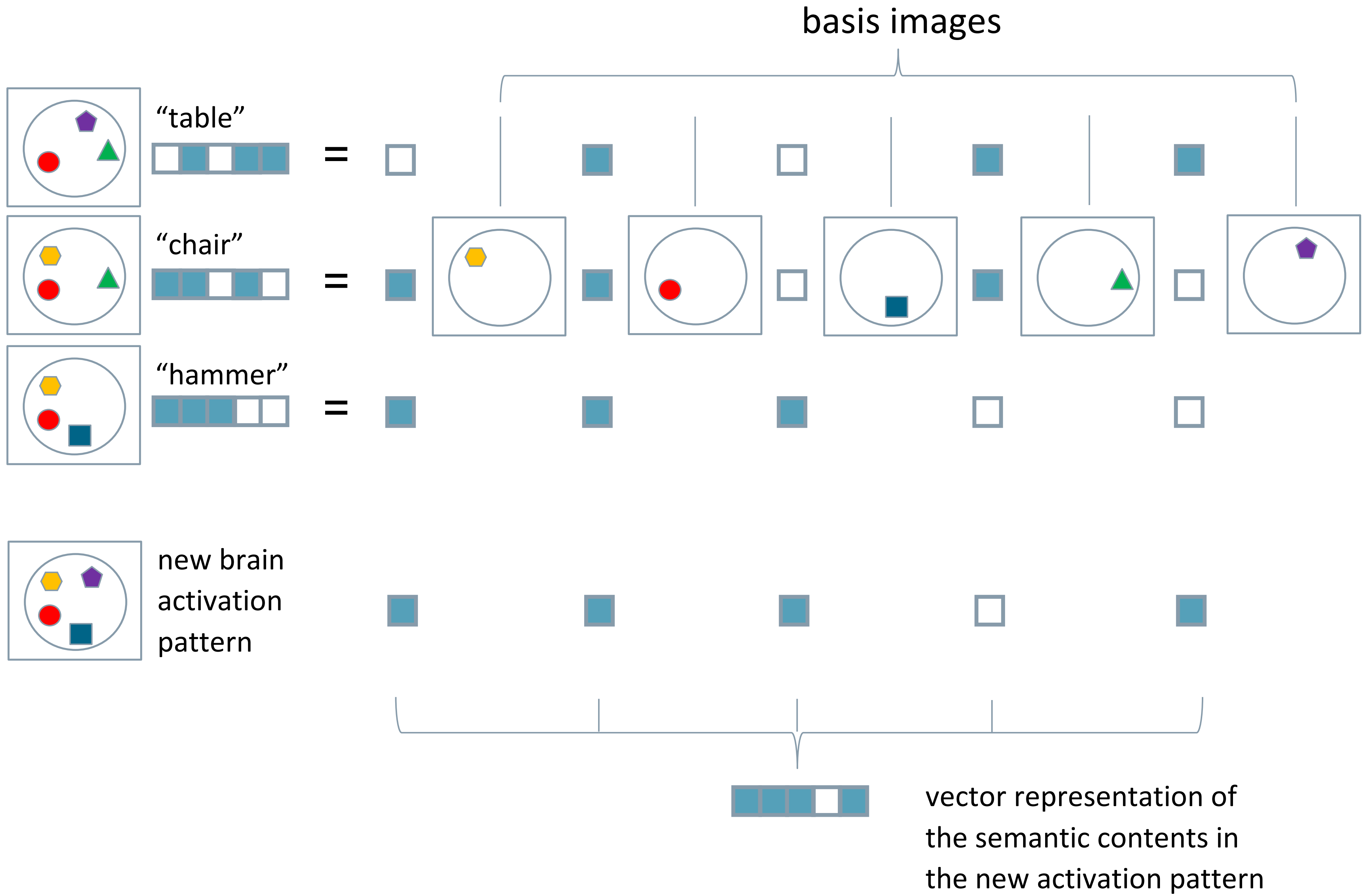
from encoding to decoding





new brain activation pattern

from encoding to decoding



case study 2 (encoding)

Identifying natural images from human brain activity

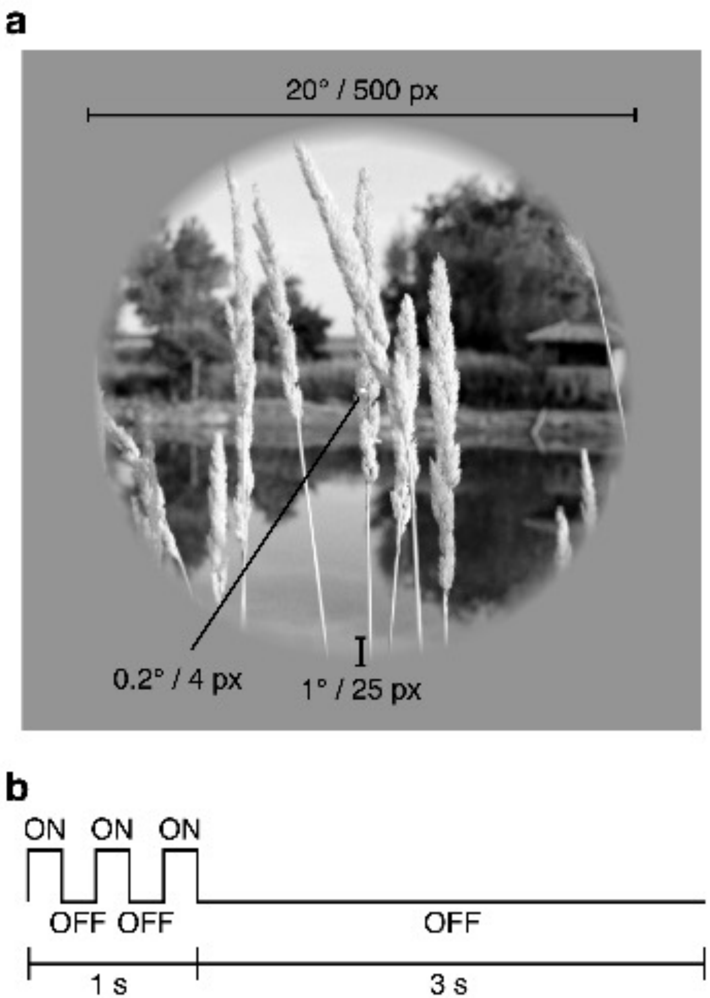
Kendrick N. Kay[1], Thomas Naselaris[2], Ryan J. Prenger[3] & Jack L. Gallant[1,2]

[Nature, 2008]

design

- 1750 training pictures
- 120 testing pictures

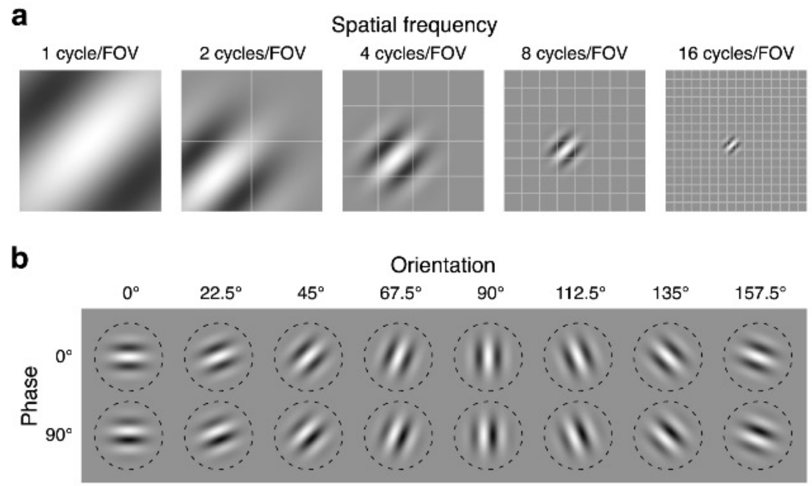
stimulus in each trial



model

representation:

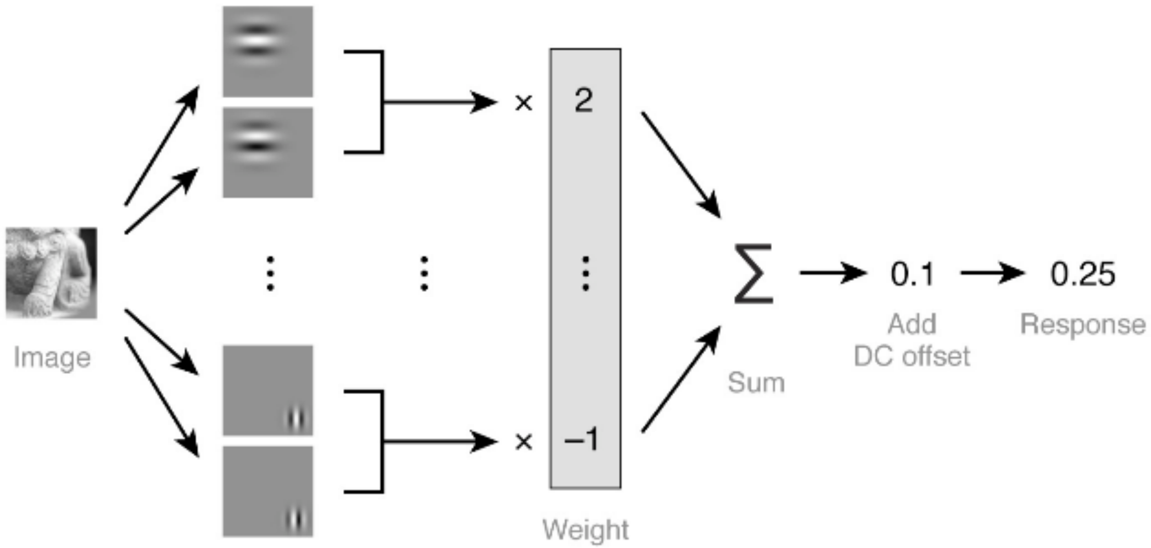
output of series of Gabor filters applied to stimulus



mapping:

each voxel is a linear combination of filter outputs

filter outputs   linear combination



evaluation

- derive representation for 120 test image stimuli
- predict activation using voxelwise mapping
- classify by similarity of predicted activation to actual activation
- accuracy out of 120 possibilities (82% on average trial data)

case study 2 (decoding)

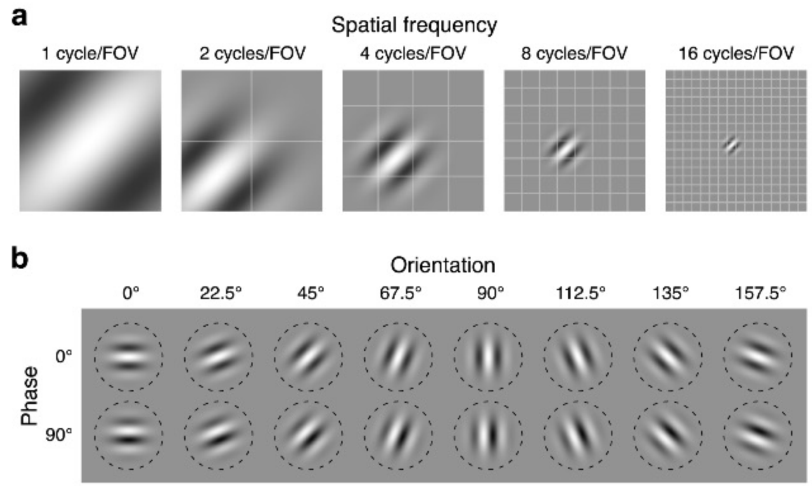
Bayesian Reconstruction of Natural Images from Human Brain Activity

Thomas Naselaris,[1] Ryan J. Prenger,[2] Kendrick N. Kay,[3] Michael Oliver,[4] and Jack L. Gallant[1,3,4,*]

[Neuron, 2009]

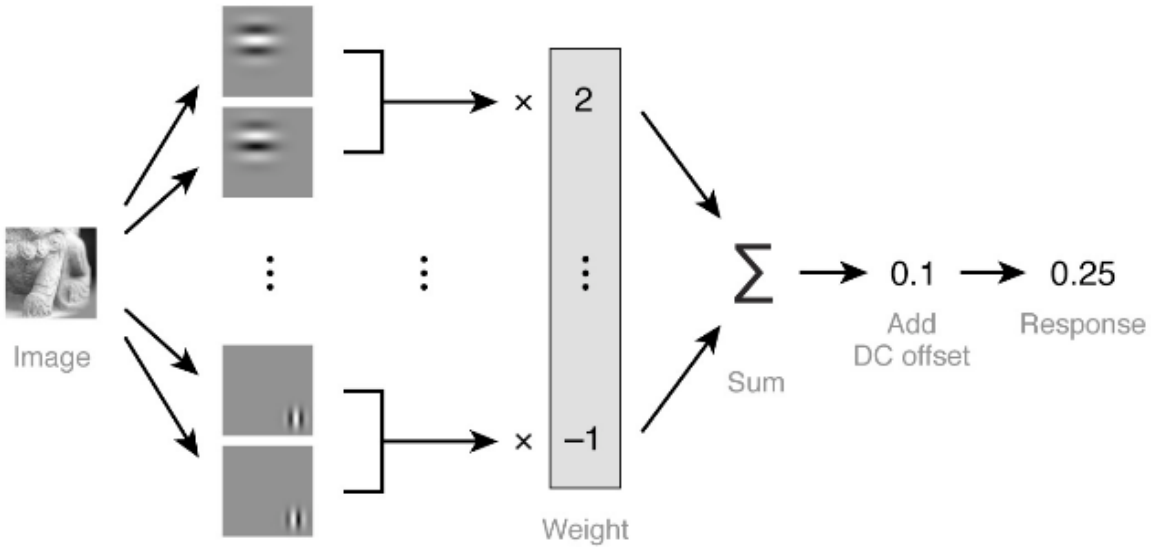
model

representation:
output of series
of Gabor filters
applied to stimulus



mapping:
each voxel is a
linear combination
of filter outputs

filter outputs linear combination



expanded model

representation:

semantic category labels for each stimulus image

mostly animate			
	human		
		many	(crowd/gathering)
		few	(body parts/portrait)
	animal		
		mammal	(land/water)
		non-mammal	(bird/fish/other)
mostly inanimate			
	man-made		
		non-building	(vehicle/artifacts)
		building	(indoor/outdoor)
	natural		
		plant	(edible/non-edible)
		non-plant	(land/water/sky)
	texture		

mapping:

each voxel is predicted as a function of semantic category

evaluation

stimulus

- invert the model that predicts each voxel as function of visual or semantic information

evaluation

- invert the model that predicts each voxel as function of visual or semantic information
- apply it to the activation data for each test stimulus:
  - obtain posterior probability for each image in a large database (millions)
  - reconstruction is the highest probability image

stimulus

evaluation

- invert the model that predicts each voxel as function of visual or semantic information
- apply it to the activation data for each test stimulus:
  - obtain posterior probability for each image in a large database (millions)
  - reconstruction is the highest probability image
- quantitative evaluation
  - correct if semantic category of reconstruction matches that of stimulus (40% on average)

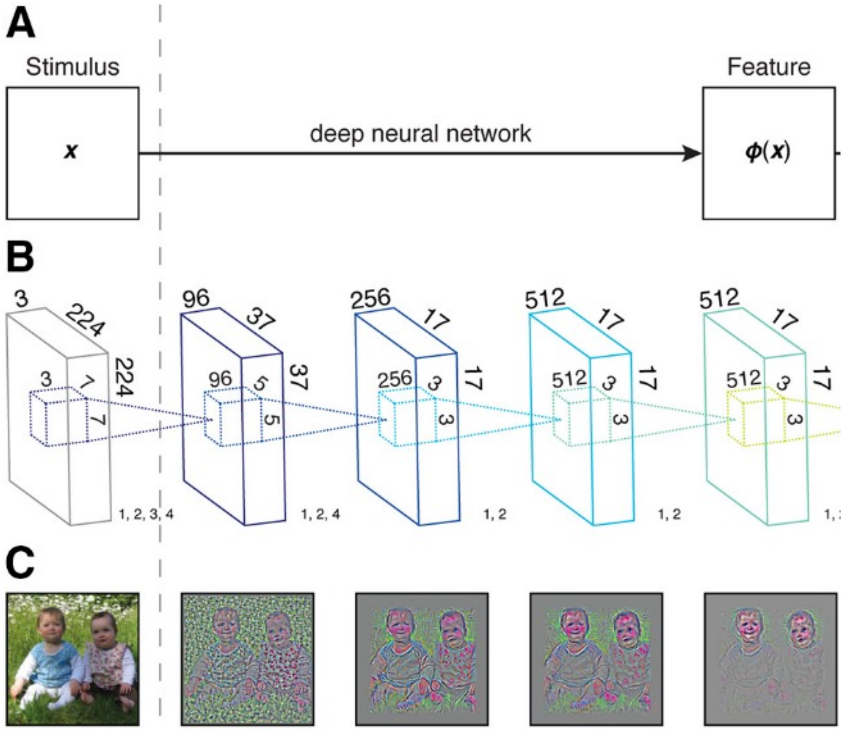
case study 2 (encoding redux)

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Umut Güçlü and Marcel A. J. van Gerven

[J. Neuro, 2015]

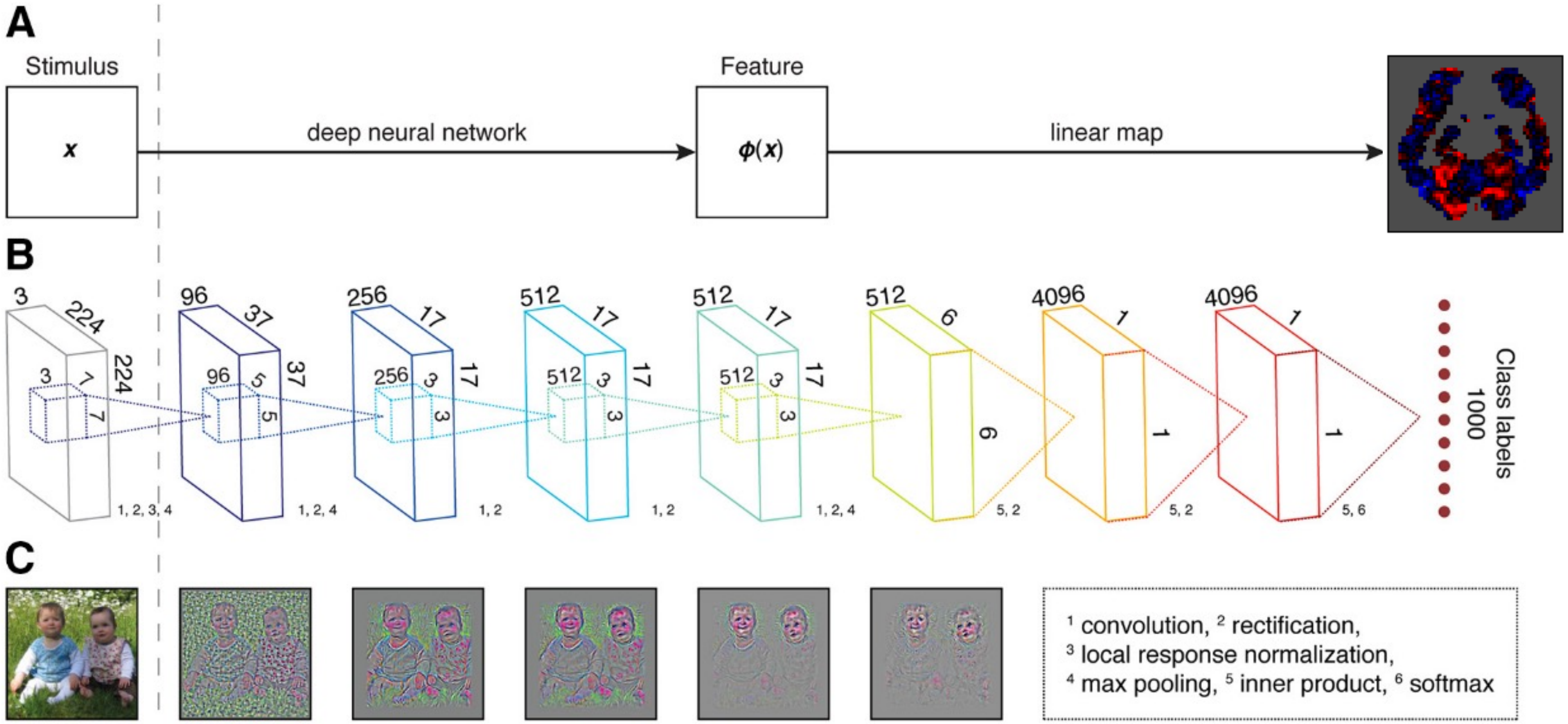
model



representation:

layers in a convolutional neural network

model



representation:

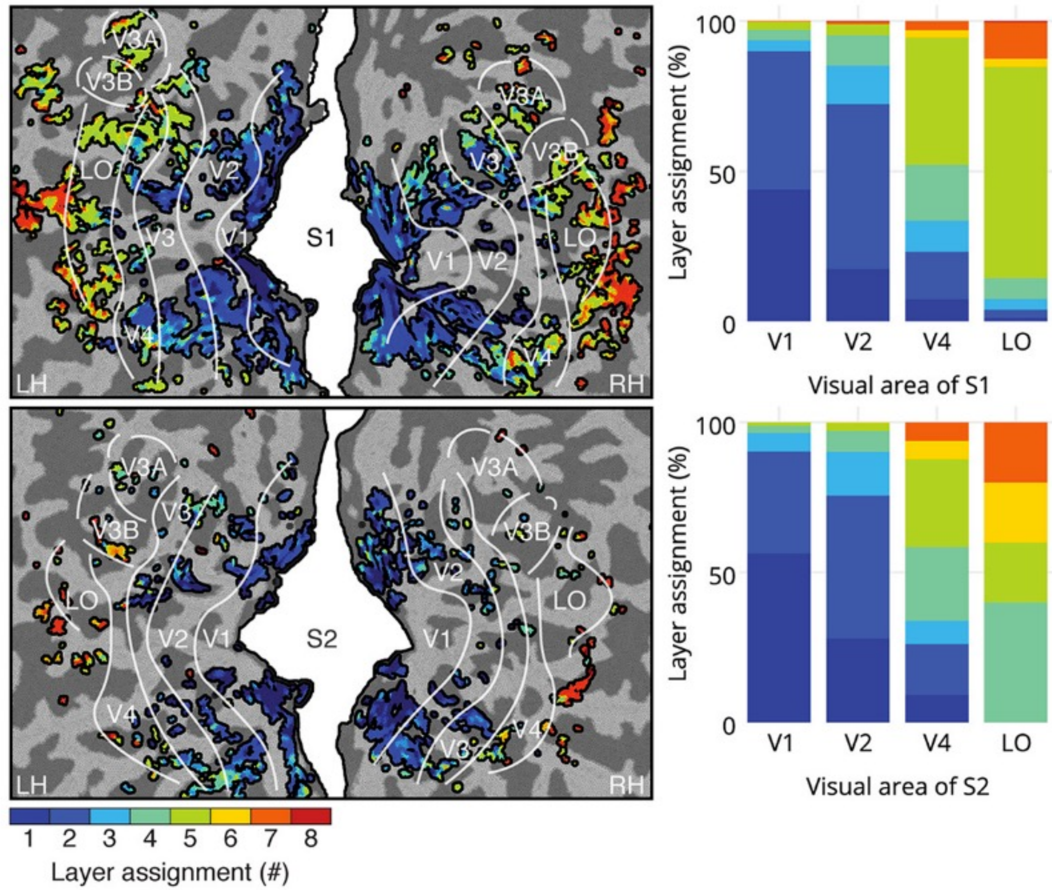
layers in a convolutional neural network

mapping:

each voxel is a linear combination of network outputs

evaluation

- assign each voxel to the network layer that best predicts it in test stimuli
- voxels that are further in the ventral visual stream are better predicted by inner network layers



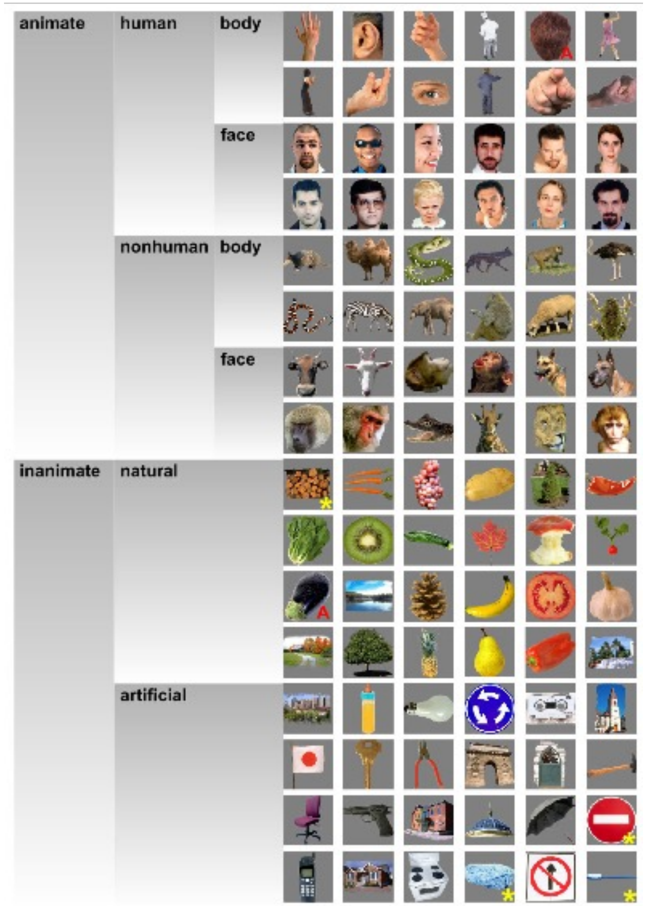
case study 3 (RSA redux)

Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation

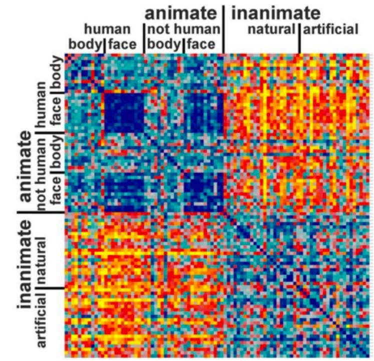
Seyed-Mahdi Khaligh-Razavi*, Nikolaus Kriegeskorte*

[PLoS Comp Bio, 2015]

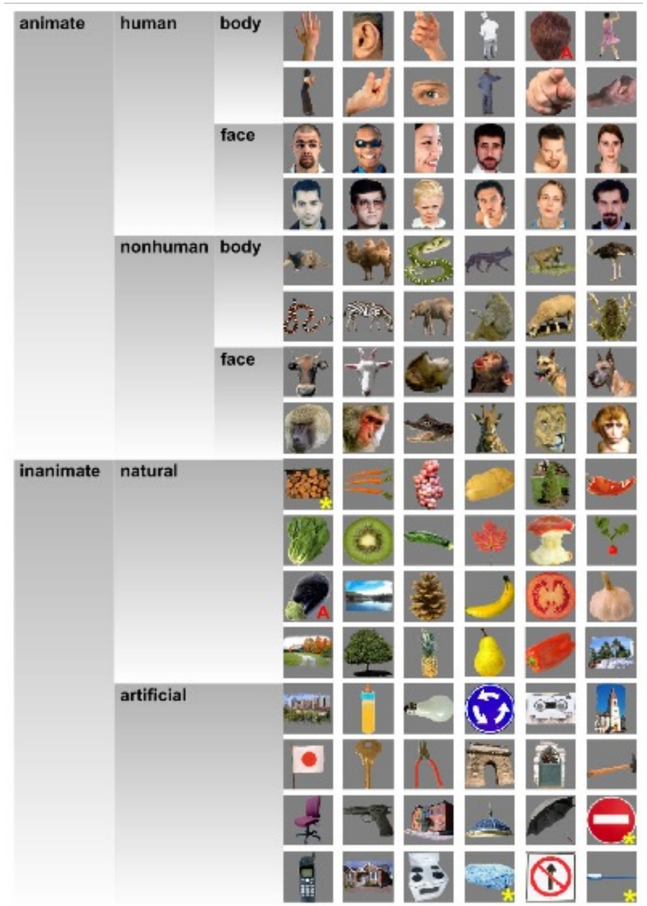
case study 3 (RSA redux)



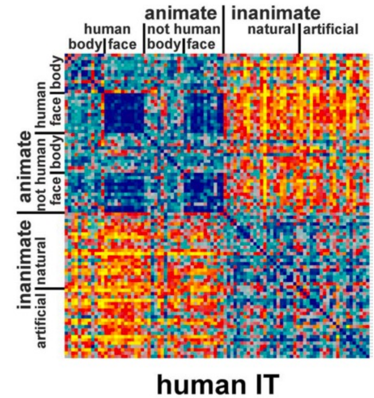
similarity of human IT activation across stimuli



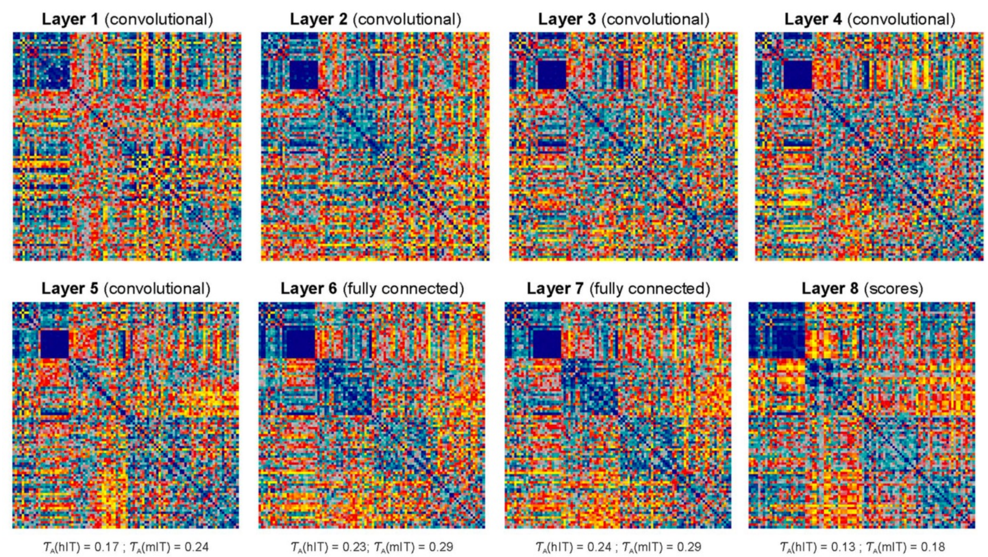
case study 3 (RSA redux)



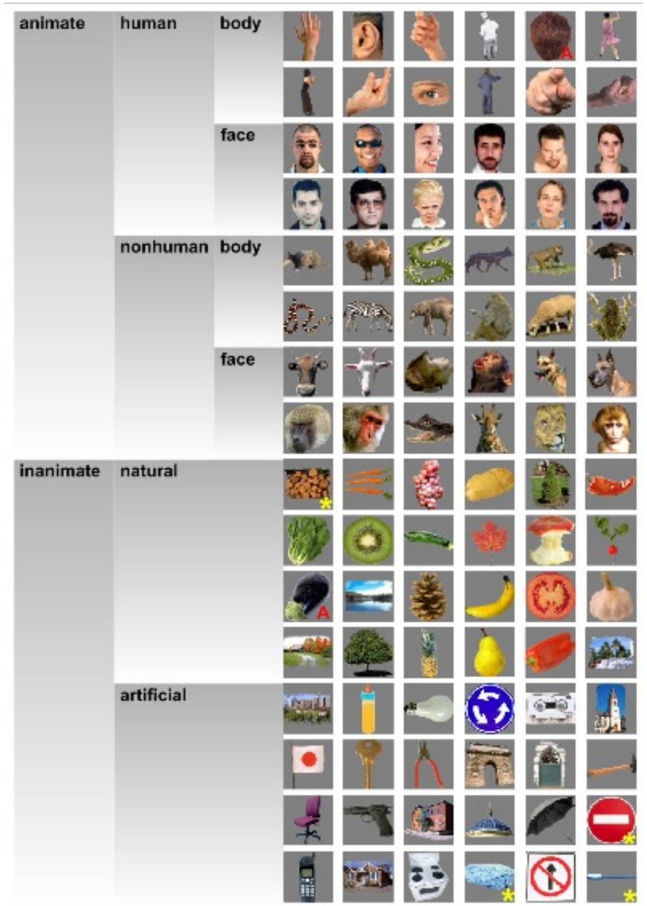
similarity of human IT activation across stimuli



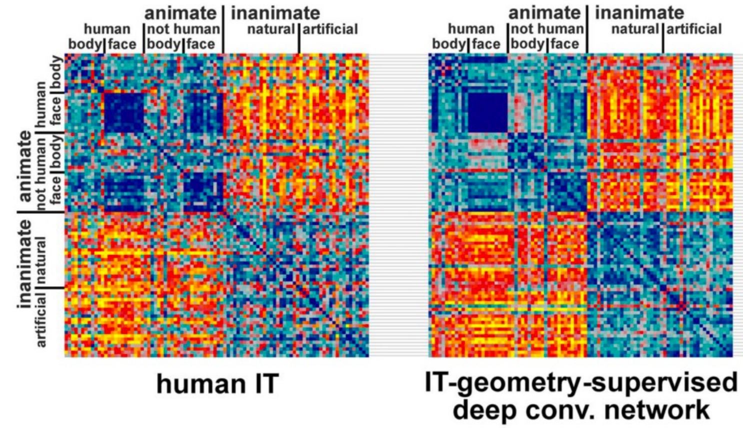
similarity of stimulus image representation in each layer of a convolutional neural network



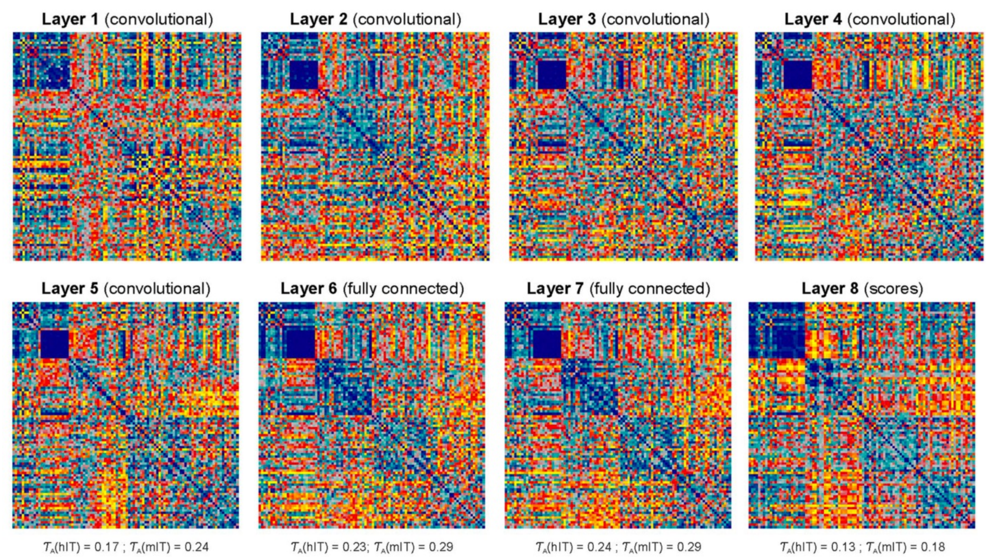
case study 3 (RSA redux)



similarity of
human IT
activation
across stimuli



similarity of stimulus image representation
in each layer of a convolutional neural network



studies based on encoding/decoding models

encoding	stimuli
Thirion 2006	binary figures
Miyawaki 2008	binary figures
Kay 2008	natural images
Mitchell 2008	word+drawing
Naselaris 2009	natural images
Just 2010	words
Nishimoto 2011	movie clips
Huth 2012	movie clips
Wehbe 2014	story (text)
Güçlü 2015	natural images
Huth 2016	story (audio)
Handjaras 2016	words (audio/text)
Anderson 2016	sentences
Anderson 2017	words
Wang 2017	sentences
Liu 2017	movies/images
...	

decoding	stimuli
Naselaris 2009	natural images
van Gerven 2010	digits
Palatucci 2011	word+drawing
Pereira 2011	word+drawing
Horikawa 2017	natural images
Liu 2017	movies/images
Pereira 2018	sentences
...	

representation similarity

- Kriegeskorte 2008
- Khaligh-Razavi 2014
- ...

summary of encoding and decoding models

- the representation is usually complex (e.g. a vector of values)
- derived from text corpora, large databases of images, behavior,...
- the same representation can be used in either direction

summary of encoding and decoding models

- the representation is usually complex (e.g. a vector of values)
- derived from text corpora, large databases of images, behavior,...
- the same representation can be used in either direction

- learn mappings from representation + imaging of training stimuli
- evaluation relies on generalization to new stimuli
  - predict imaging data or infer representation
  - in the limit, actual reconstruction of the stimulus!
  - prior information helps (what could it be, statistics of natural images, etc)

summary of encoding and decoding models

encoding

- identify voxels/locations the model can predict
- classify predicted activation by similarity with true activation

decoding

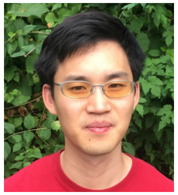
- extract the representation from activation for novel stimuli
- reconstruct stimulus or an approximation thereof

representation similarity

- can be done in either encoding or decoding model
- compare either activation or representation similarity
  with reference similarities obtained in various ways

the machine learning team

Francisco Pereira



Charles Zheng

Patrick McClure

we can help with

- turning stimuli into representations (automatically, if we are lucky!)
- deriving representations from behavior or other sources
- devising an encoding/decoding model strategy for your problem...
- ... or using all the methods described earlier...

email francisco.pereira@nih.gov or drop by (B10, 3D41)

Thank you!